

DRAFT (CHAPTER 2)

Inverse Modeling and Data Assimilation in Geosciences: A Practical Guide

Author : Dr. Tomislava Vukicevic

Affiliation : ATOC University of Colorado

Date of last revision: May 15 2006

2. Inverse problem basics

2.1 Parameter and model space

Let a model within scope of the Geoscience problems be denoted $\phi(m)$ with control parameters m , where m is a set of physical quantities. For each choice of parameter values (quantitative specification of the parameters in m) there is different realization or simulation with ϕ . The space or manifold spanned by the different values of m is called ***parameter space***. The parameter space is populated with possible values of the parameters. The quantitative results of ϕ for different choices of the parameter values also span a space or manifold, called the ***model space***. The existence of space of ϕ is interpreted as the existence of different possible simulations of measurements by the model which represents the same governing laws.

There are two possible distinct causes of the existence of parameter space. First, the parameter values are characterized with errors around some known reference set of

values. Second, the different parameter values are result of variability by natural causes of the physical system which is modeled by ϕ in the domain of characterization by the control parameters. This implies that the governing laws are well known but the processes which determine quantification of the parameters are not. These two interpretations and the consequence of them are illustrated in the following example

Example 1.1: Damped oscillations

The phenomenon of interest is evolution of displacements of a mass from equilibrium points under the influence of a potential field or elasticity force and a dissipation force. The simplest physical model of this kind is a mass on a spring as in Figure 1.1

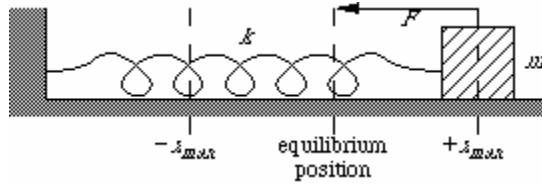


Figure 1.1: Mass on a spring

The governing equation for evolution of the displacement over time is written

$$\frac{d^2\chi}{d^2\tau} + \alpha \frac{d\chi}{d\tau} + \omega\chi = 0 \quad (1.1)$$

where χ is the displacement from the equilibrium point (also called oscillation amplitude), τ is time, α is damping coefficient which represents dissipative force such as friction and air resistance and ω is spring elasticity coefficient which determines

frequency of oscillations. The equation for the model in Figure 1.1 is derived directly from application of the Second Newton's law and the Hooke's law (1676) to which the dissipation force is added.

The oscillator model approximates many natural systems that vibrate or oscillate. In general it is used to represent governing equations for a particle moving through any region whose potential has one or more local minima: pendulum, planetary and satellite motion, the classical description of an electron in orbit around a nucleus and an air parcel in geopotential field, to mention some. The solution of (1.1) is readily obtained

$$\begin{aligned}
 \chi &= \exp(r\tau) \\
 r^2 + \alpha r + \omega &= 0 \\
 \lambda &\equiv \frac{\alpha}{2} \\
 \omega_0^2 &\equiv \omega \\
 r &= -\lambda \pm \sqrt{\lambda^2 - \omega_0^2} = -\lambda \pm \eta \\
 \chi &= A_1 e^{(-\lambda+\eta)\tau} + A_2 e^{(-\lambda-\eta)\tau}
 \end{aligned} \tag{1.2}$$

The damped oscillation solution depends on two initial conditions (arbitrary starting time with $\tau = 0$), and values of α and ω . These are the control parameters in the model (1.1). There are many, in fact infinite, possible values which could be assigned to the control parameters. Samples of the solution (1.2) are shown in Figure 1.2. The figure illustrates that varying of the initial conditions causes change in the amplitude of χ but not frequency, while the variability of α and ω results in the change of both the amplitude and frequency. Obviously, in the asymptotic limit $\tau \rightarrow \infty$, the solution is the state at rest ($\chi = 0$), implying that any values in the set $[\alpha, \omega, A_1, A_2]$ would produce very similar quantification of the state in the long time limit.

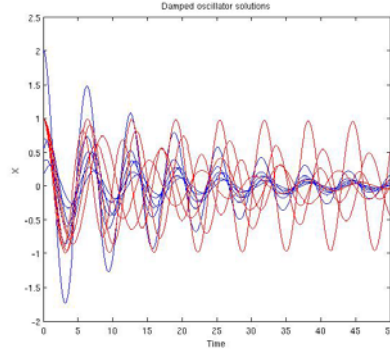


Figure 1.1 Sample of solution (1.2) for different control parameter values

The interest is typically in the oscillation amplitudes at arbitrary but finite times, as would be in many other problems where an evolving natural system is modeled. A choice of the reference values of $[\alpha, \omega, A_1, A_2]$ with errors implies that the forces and initial conditions are pretty much known. The example of oscillator model space spanned by simulations from the parameter space with small errors around the reference is shown in Figure 1.2-A. Second choice is the parameter space spanned by values of α and ω from different elasticity of the spring or different medium which exerts the resistance. The range of possible values of α and ω must be larger than in the simulations with errors around the reference. The resulting model space is shown in Figure 1.2-B. Comparison indicates that the range in the model space increases with increasing range of the possible parameter values, as expected, but the two spaces are not very different in the mean at early times (panels a and d). The model simulations at larger times show tendency to cluster closer to zero for the simulations shown in 1.2-B, indicating that there are many combinations of α and ω values which result in the similar small values in the model space. This example shows that the model space is not uniquely defined by the

origin of parameter space, consistent with common wisdom that there are potentially many possible causes resulting in the same measurable quantity.

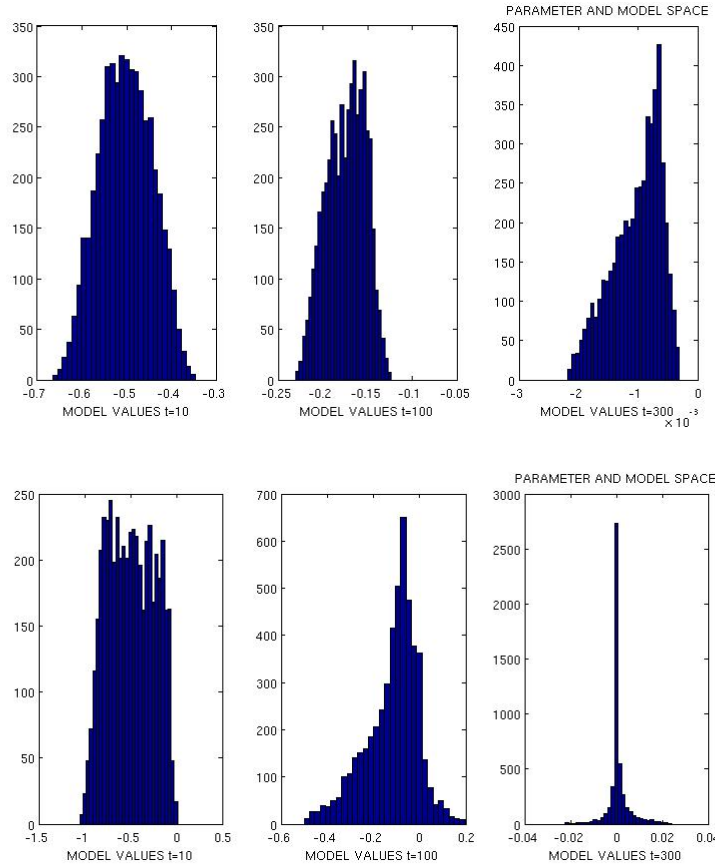


Figure 1.2 Histogram of simulated oscillation amplitudes at 3 time points (left to right) which resulted from parameter space of α and ω with: A – small range , and B - large rang. t

2.2 Measurement space

The measurement space is simpler to define than the parameter space. It is the space spanned by possible values of the measured quantity within the uncertainty range of the measuring procedure. The measuring procedure could include multiple measurements of the same quantity or multiple measurements of different quantities but

for the same realization of the natural system. In the latter case the measurements constitute a multidimensional space similar to the control parameter set. In the oscillator model (1.1) there is only one dimensional phase space represented by χ . The measurable quantity in this example is $\chi(\tau_i)$, where τ_i is discrete time.

In the inverse problem there is explicitly derived dependency of the parameter to measurement uncertainty which is presented in the next section. Here the interest is to discuss the consequence of existence of the measurement space spanned by the measurement uncertainties. The range of control parameter values which would result from the measurement uncertainty is interpreted as in the forward problem as the range of uncertainty on the parameters. This property emphasizes the critical property of the modelization of the natural systems: ***When it is necessary to solve the inverse problem in the process of understanding and modeling of the natural system, the uncertainties in the measurements would render the uncertainties in estimates of what controls the system as hypothesized by the system model.***

The parameter space which results from the variable external causes leading to the variable parameter values is related to the measurement space in more complex way than the measurement uncertainties. Each individual measurement is a recorded quantity of a response of an instrument to the medium that is measured. The medium when measured is at one specific state after one realization of the possible external cause. In order to capture natural variability of the parameters in the inverse problem solution which is caused by conditions external to the model, it is necessary to evaluate it from many measurements and different state realizations. It is shown later that validity of an

evaluated range of actual variability in the inverse problem solution would depend on three factors: 1) abundance of measurements, 2) size of measurement errors and 3) strength of sensitivity of the forward model to the control parameters. Analysis of impact of each of these factors is important subject in specific applications as it addresses potential to distinguish different causes of the natural phenomena by the specific model and available measurements.

2.3 Probabilistic nature of information in the inverse problem

The property of measurements to always have errors makes them random quantities. Consequently, the model control parameters which would be derived from the inverse problem solution using the measurements would also be random quantities. Even without the inverse problem the model control parameters could be random quantities if their values are uncertain. The random quantity, also called the stochastic quantity, is a quantity which exact value is not known or predictable. What is known about the random quantity is a possible value from a range with an associated probability. Because the measurements and control parameters are by design the stochastic quantities, relationships between these quantities in the inverse modeling problem and applications in the data assimilation problems must be derived based on the relationship between the associated probabilities.

2.3.1 Interpretation of probability

First, let A be realization of a stochastic physical quantity with the numerical value from within an interval $(x, x + dx)$. If there are many realizations of A , it would be possible to derive probability of A as chance of occurrence of A . A is then an event with probability $P(A)$ for which the following classical axioms of probability apply

$$\begin{aligned} P(A) &> 0 \\ P(\emptyset) &= 0 \end{aligned} \tag{1.3}$$

If A and B are disjoint events

$$P(A \cup B) = P(A) + P(B) \tag{1.4}$$

If A and B are not disjoint events

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.5}$$

where $P(A \cap B)$ is joint probability

A distribution of probabilities over the space of possible values of A defines the probability distribution on that space. Another important function associated with the

stochastic quantity and its probability distribution is the probability density $p(x)$ which satisfies

$$P(A) = \int_A p(x)dx \quad (1.6)$$

where x represents coordinates, indicating that in the general case, the event A is a set of physical quantities included in A , such as the set of control parameters or set of measurements. The probability density function is of critical importance in the description of the stochastic quantities because when it is known they are completely described.

There is another intuitive way to interpret the probability of stochastic physical quantity in the inverse modeling problem. The probability could be defined as in Tarantola (2005) as : “ *subjective degree of knowledge of the true value*”. It is somewhat difficult to understand the emphasis on subjective knowledge in the Tranatola’s definition, but it is instructive to consider the interpretation of probability which uses the reference to the truth. In this approach the uncertainty or error which renders the physical quantities stochastic is measured as deviation from the truth. It is shown later that even when the truth is not known, which is most of the time, the uncertainty defined as deviation from the truth could be sensible approach to interpreting the probability in the results of the inverse modeling problems. The probabilistic variables and relationships (1.3 – 1.6) are the same for either interpretation of the probability.

2.4 General inverse problem and solution

2.4.1 Conditional probability

The key relationship which links the probabilities of stochastic quantities in the problem of evaluating the control parameters by inversion from the measurements is most commonly expressed by the Bayes' rule (1763) for conditional probabilities.

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} \quad (1.7)$$

where A and B are statistical events. The rule is actually derived from the definition of conditional probability

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (1.8)$$

The left hand side is read as “probability of A given B”. From the definition it follows

$$\begin{aligned} P(A \cap B) &= P(A/B)P(B) \\ P(A \cap B) &= P(B/A)P(A) \end{aligned} \quad (1.9)$$

The Bayes' rule is then readily derived

Assuming that the event B is from the control parameter space and A from the measurement space, then the rule (1.7) is read as:

“Conditional probability of the parameter taking values defined by the event B conditioned on the measurement taking values as defined by the event A is equal to product of conditional probability of the measurement taking values defined by the event A conditioned on the parameter taking values as defined by the event B and probability of the parameter taking the values as defined by the event B , normalized by probability of the measurement taking values as defined by the event A ”

This relationship apparently allows to evaluate probability of the parameter (as defined by B) given the measurements (as defined by A) assuming that right hand side (r.h.s.) of (1.7) is known. The probabilities $P(A/B)$, $P(A)$ and $P(B)$ are hard to evaluate when based on the occurrence of events approach. It is far more convenient to assume probability distributions associated with the space to which the events A and B belong (i.e., the measurement and parameter spaces, respectively). As the distribution is determined by the probability density (1.6), the problem is then transformed to finding a relationship between the probability densities on the joint parameter and measurement spaces. To arrive at the relationship which relates the probability densities instead of the probabilities of individual events we take the approach from Tarantola (2005) of defining the probability densities in the joint spaces of the parameters and measurements and their conjunction .

It is possible but not necessary to derive the desired relationship between the probability densities as generalization of the Bayes' rule (1.7). This approach is taken in the literature on estimation and stochasting filtering theory which addresses the inference of state of modeled time evolving systems from discrete stochastic measurements (Jazwinski, 1970; Sorenson, 1985). In the applications in the Geoscience problem examples of the use of equivalent to the Bayes' rule for probabilities is described in Cohn (1997), Rodgers (2000), [\(more references\)](#). In the stochastic filtering theory literature the generalization of the Bayes' rule is derived by a limiting process in the joint space of the measurements and modeled state (Jazwinski, 1970). It is beyond the scope of this text to present the theoretical derivation and indebt analysis of the use of conditional instead of the joint posterior probability density functions. In the present chapter the approach from Tarantola (2005) is adapted for easy interpretation of the origin of probability density functions on the parameter and measurement spaces which apply within wide scope of the Geoscience problems where the parameters and models of many kinds are used to analyze and predict the state in conjunction with vast variety of measurements.

2.4.2 Conjunction of probability distributions

It is shown in section 2.1 that there are two sources of information about the natural system under study. These are the modeled and measured information. Let parameter space be denoted M , spanned by points (m_1, m_2, \dots) . This space is transformed into a measurement space by a forward model

$$d = \phi(m) \tag{1.10}$$

In the damped oscillator example ϕ is represented by equation (1.1). Let the measurement space as simulated by the model be denoted O . O is spanned by points (d_1, d_2, \dots) . The joint space $\Omega = M \times O$, which is characterized by a joint probability density $f(m, d)$, is the space of all possible information available about a natural system under study, given the model. The joint probability density on Ω provides complete description of the uncertainties and natural variability in the parameters and the result of these by the model simulations which is contained in the space O . The joint probability density $f(m, d)$ could also include effects of modeling errors. The modeling errors would result from the use of imperfect model. For example, the damped oscillator model (1.1) may be used to simulate damped oscillations which are in reality also driven by some unknown external harmonic force. When the force is not included in the equation, the model would be in error relative to the actual natural system and consequently relative to the measurements. It is not trivial task to design or assume the effect of modeling errors when specifying the joint probability density $f(m, d)$. This problem is illustrated in the exercises ??.

The other information about the natural system is contained in the actual measurements which are independent of the model. Let this information be in space denoted C . There is a joint probability density on the joint space $\Theta = C \times M$, denoted $\rho(m, d)$. Notice that $\rho(m, d) \neq f(m, d)$. The union of measurement spaces O and C defines total measurement space which is denoted D . The joint probability densities $f(m, d)$ and $\rho(m, d)$ are both defined on D . New information about the system would

be obtained when the two joint probability densities are combined by conjunction (Tarantola, 2005, chapter 1.5)

$$p(m,d) = \frac{1}{\gamma} \frac{\rho(m,d)f(m,d)}{v(m,d)} \quad (1.11)$$

Where $\gamma = \int_{D \times M} \frac{\rho(m,d)f(m,d)}{v(m,d)}$ is constant and v is so called homogenous probability density. $p(m,d)$ is *a posteriori* probability density on the joint space $D \times M$ resulting from the combined probability distributions. The knowledge of a posteriori probability density is the most complete available quantitative knowledge of information about the natural system under study. By this property, the expression (1.11) defines the general inverse modeling problem:

Evaluate $p(m,d)$ from knowledge of $\rho(m,d)$, $f(m,d)$ and $v(m,d)$.

$p(m,d)$ contains all available quantitative information of the system in the space $M \times D$ from which solution of the inverse modeling problem is to be derived. To arrive at the resolution we need to introduce definitions of marginal and conditional probability densities and a priori information.

The marginal probability densities for any joint space with the associated joint distribution $g(a,b)$ in the space spanned by points $(a_1, a_2, \dots, b_1, b_2, \dots)$ are

$$\begin{aligned} g_A(a) &= \int_B g(a,b)db \\ g_B(b) &= \int_A g(a,b)da \end{aligned} \quad (1.12)$$

When the space (a_1, a_2, \dots) is independent of (b_1, b_2, \dots) then

$$g(a,b) = g_A(a)g_B(b) \quad (1.13)$$

The conditional probability density is defined

$$g_{A/B}(a/b) = \frac{g(a,b)}{g_B(b)} \quad (1.14)$$

The conditional probability density is interpreted as the probability density of points in the joint space for which $b = b(a)$. Using (1.14) the joint probability density for the information given the model is

$$f(m,d) = f(d/m)v(m) \quad (1.15)$$

where the marginal probability density in the parameter space is assumed to be equal to the homogenous probability density of the parameters. The conditional probability density $f(d/m)$ is made of the results of forward model applied over a space of control parameters **without knowledge of the measurements**. In figure 1.2 the discrete examples of this probability density are shown for the damped oscillator model.

The probability density $\rho(m,d)$ results from the information in the joint space of the control parameters and measurements without knowledge of the model. It is natural to assume that these are independent. From (1.13)

$$\rho(m, d) = \rho_M(m) \rho_D(d) \quad (1.16)$$

The probability density $\rho_D(d)$ results exclusively from information about the uncertainties in the measurements. The probability density $\rho_M(m)$ in turn results from information of the uncertainties or variability in the control parameters which is independent of the measurements. This information is called *a priori*.

Under the same assumption as in (1.16) the homogenous probability density in (1.11) is

$$\nu(m, d) = \nu_M(m) \nu_D(d) \quad (1.17)$$

Substituting (1.15-1.17) into (1.11) renders

$$p(m, d) = \frac{1}{\gamma} \frac{\rho_D(d) \rho_M(m) f(d / m)}{\nu_D(d)} \quad (1.18)$$

The solution of the general problem (1.18) is to compute the marginal probability density for the control parameter space. Using (1.18) in (1.12)

$$p_M(m) = \int_D \frac{1}{\gamma} \frac{\rho_D(d) \rho_M(m) f(d / m)}{\nu_D(d)} D d \quad (1.19)$$

(Tarantola, 2005). $p_M(m)$ is interpreted as projection onto M . The probability densities on the r.h.s. of (1.19) are assumed known for the specific application. Geosciences and other physical sciences where parameters and models of many kinds are used to analyze and predict the state in conjunction with vast variety of measurements.

In the present definition of the joint space $M \times D$ with the associated joint probability density $p(m, d)$, the conditional probability density in for the parameter space could be derived from application of (1.14) assuming existence of $m = m(d)$. This assumption is somewhat difficult to interpret in the general case in which the parameter space is not the same as the modeled system state as in the stochastic filtering theory. When the assumption is valid it implies

$$p(m, d) = p(m/d)p_D(d) \quad (1.20)$$

Combining (1.18) and (1.20)

$$p(m/d) = \frac{1}{\gamma} \frac{\rho_M(m)\rho_D(d)f(d/m)}{v_D(d)p_D(d)} \quad (1.21)$$

This expression implies that the conditional probability density of parameters conditioned on the measurements is obtainable from the independent information about quantities in the space of measurements and parameters. The posterior probability densities in (1.19) and (1.21) are apparently different, but in either case the required knowledge about the independent stochastic information in the parameter and

measurements spaces is the same. Before addressing common choices in general and more specifically in the examples in chapters 4 and 5 it is instructive to consider what type of information may be most useful or interesting to derive from the knowledge of posterior probability density function.

2.4.3 Estimation criteria

In the practice with the Geoscience problems the parameter space is often large multidimensional space. In this situation it is unfeasible to either evaluate or visualize (1.19) or (1.22). Instead, characteristics of the posterior probability density function are used to define single *best estimate* or *central estimate* of the parameters (Cohn, 1997, Jazwinski 1970; Tarantola, 2005). The commonly used central *estimation criteria* are

- a) **Maximum likelihood**, define by a discrete region or continuous point with maximum probability associated with the posterior probability density function. The likelihood function is

$$L(m) = \int_D \frac{\rho_D(d)f(d/m)}{\nu_D(d)} dd$$

- b) **Minimum variance**, defined by the mean of the posterior probability distribution

$$\langle m \rangle = \int_M mp_M(m)dm \text{ or conditional mean } \langle m \rangle = \int_M mp(m/d)dm$$

- c) **Minimum absolute distance**, defined by the median of the posterior density distribution

The choice of criterion would depend on the purpose of estimation and characteristics of the specific problem. For example,

2.4.4 Conjunction of Gaussian distributions

It is common to assume that probability density functions associated with model and measurement spaces are Gaussian. The Gaussian distribution is characterized with only two statistical parameters: mean $\langle x \rangle$ and covariance C

$$p(x) = \frac{1}{(2\pi)^{\frac{1}{2}} \det^{\frac{1}{2}} C} \exp\left(-\frac{1}{2} (x - \langle x \rangle)^T C^{-1} (x - \langle x \rangle)\right) \quad (1.23)$$

In the model space (1.23) is

$$f(d / m) = \frac{1}{(2\pi)^{\frac{1}{2}} \det|C_s|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (d - \phi(m))^T C_s^{-1} (d - \phi(m))\right) \quad (1.24)$$

while in the measurement space

$$\rho_D(d) = \frac{1}{(2\pi)^{\frac{1}{2}} \det|C_d|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (d - d_{meas})^T C_d^{-1} (d - d_{meas})\right) \quad (1.25)$$

Using (1.24) and (1.25) in (1.19)

$$p_M(m) = k \rho_M(m) \exp\left(-\frac{1}{2} (\phi(m) - d_{meas})^T C_D^{-1} (\phi(m) - d_{meas})\right) \quad (1.26)$$

Where d_{meas} denotes actual measurements, k is cumulative constant and

$$C_D = C_d + C_s \quad (1.27)$$

(1.26) indicates that the conjunction of two Gaussian probability distributions in the measurement space is also Gaussian with the summed up uncertainties from the independent modeled and measured information, represented by the covariance C_D

Problem 1: Derive 1.27 (Appendix)

When it is further assumed that the a priori probability density function in the parameter space is Gaussian

$$\rho_M(m) = \frac{1}{(2\pi)^{\frac{1}{2}} \det|C_m|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(m - m_{prior})^T C_m^{-1} (m - m_{prior})\right) \quad (1.28)$$

then

$$p_M(m) = \text{const} \exp(-S(m))$$

$$S(m) = \frac{1}{2} \left[(\phi(m) - d_{meas})^T C_D^{-1} (\phi(m) - d_{meas}) + (m - m_{prior})^T C_m^{-1} (m - m_{prior}) \right]$$

$$S(x(0)) = \frac{1}{2} \sum_{time} \left(H(x(\tau) - d_{meas}(\tau))^T C_D^{-1} (H(x(\tau) - d_{meas}(\tau)) + (x(0) - x_{guess}(0))^T C_{guess}^{-1} (x(0) - x_{guess}(0)) \right) \quad (1.29)$$

$S(m)$ is apparently weighted sum of squares. When the model is linear $\phi(m) \equiv Fm$, then

$p_M(m)$ in (1.29) becomes Gaussian with the mean and covariance, respectively

$$\begin{aligned} \langle m \rangle &= m_{prior} + C_M F^T (F C_M F^T + C_D)^{-1} (d_{meas} - F m_{prior}) \\ C_S &= (F^T C_D^{-1} F + C_M^{-1})^{-1} \end{aligned} \quad (1.30)$$

Problem 2: Derive 1.30 (Appendix)

In the section on Kalman Filter technique (3.2) it is shown that solution (1.30) is also derived for the data assimilation problem by the stochasting filtering theory which

addresses the inference of state of modeled time evolving systems from discrete stochastic measurements (Jazwinski, 1970). This theory is applicable in the Atmospheric sciences and Oceanography when the interest is to produce quantification of the atmospheric or oceanic state in geographical discretized space and over time (Cohn, 1997; Kalnay 2000).

Application of the maximum likelihood criterion for the central estimate by (1.29) implies minimization of $S(m)$. The minimization of $S(m)$ is commonly referred to as “least square problem” which is treated in the chapter on Variational techniques (3.3).